

Privacy Aspects of Health Related Information Sharing in Online Social Networks

Sadeq Torabi and Konstantin Beznosov
 Department of Electrical and Computer Engineering
 University of British Columbia
 Vancouver, BC, Canada
 {sadeq,beznosov}@ece.ubc.ca

Abstract—Online social networks (OSNs) have formed virtual social networks where people meet and share information. Among all shared information, health related information (HRI) has received considerable attention from researchers and individual users. While considered beneficial, sharing HRI, which is personal in nature, comes with its privacy drawback. Privacy is a process of boundary regulation that is related to the individual and her perception of the surrounding environment. As a result, the subjective privacy risk perceptions associated with sharing HRI in OSN have driven people to adopt different types of behaviour, both in terms of HRI sharing and privacy risk mitigation.

Through an online survey, we examined factors that affect users’ perceived privacy risks along with their risk-mitigating behaviour, when it comes to sharing HRI in OSNs. The results suggest that the majority (over 95%) of participants share some HRI, with the “type” and the “recipient” of the shared HRI being the key factors that affect the perceived privacy risk and the risk-mitigating behavioural responses.

Index Terms—HRI, OSN, Risk Perceptions, Behavioural Responses, Mechanical Turk.

I. INTRODUCTION

Sharing health related information (HRI),¹ such as symptoms, treatments, prescriptions, and diet related information, could be beneficial for individuals and their social networks. For example, social pressure has been shown to be an effective incentive for losing weight and can influence people to make healthier lifestyle choices [1, 2]. In addition, patients with serious illnesses can learn from other individuals with similar conditions by connecting through the Internet [3, 4]. Social support is also shown to be effective in maintaining physical and mental health during certain long-term medical treatments [5]. In fact, the highly inter-connected nature of existing online social networks (OSNs), and the increasing number of OSN users have encouraged people to actively use OSNs for sharing different types of information, including HRI. According to recent Pew Internet report, about 23% of online users in the US have followed their friends’ personal health experience in the past year, with a 3% increase compared to 2010 [6]. Meanwhile, 16% of the surveyed participants reported going online to find others who shared similar health concerns [7]. Other surveys have also shown that sharing health related knowledge online is in fact becoming a leading habit among people [3, 4, 6, 7].

¹We define HRI as any type of information that is related to the health of an individual or people within her social network.

Despite the popularity of the OSNs and their large number of users, existing OSNs suffer from security and privacy issues, making them vulnerable to different attacks [8–12]. Inherently, OSN users’ HRI, which is thought to be personal and private by nature, might be subject to unintended disclosure. Privacy invasion, stalking, data re-identification (de-anonymisation), medical data misuse, and damaging personal data are a few examples of privacy and security related threats associated with unintended information disclosure [11, 13, 14]. In reality, depending on their knowledge and previous experience, people perceive privacy risks associated with sharing their HRI in OSNs differently [3, 4, 13, 15–17]. Their behavioural responses to perceived privacy risks vary as well [18]. People may avoid risks by manipulating their information, or cope with the perceived risks for the sake of getting benefits, or simply avoid sharing their HRI if the perceived risks are high [18]. For example, a study by Velden and El-Emam showed that younger patients who used Facebook to share different personal information were not willing to share their HRI in Facebook and preferred to act as “regular” (i.e., with no disease) when communicating with their Facebook connections [17]. On the other hand, a study of US mothers who used online technologies to share or look for HRI showed that about 70% of them share HRI through their Facebook accounts [19].

Motivated by the need for continuous assessment of users’ subjective behaviours, we constructed an online survey to investigate existing practices of sharing HRI in OSNs. We also explored users’ privacy risks perceptions and risk-mitigating behavioural responses. We recruited 166 active OSN users through Amazon Mechanical Turk (MTurk) [20] and surveyed their sharing behaviour for different HRI categories. We analyzed characteristics of HRI instances from the literature [5–7] to group them into 8 categories (Table I). The majority of participants (95.8%), indicated sharing some HRI through their OSN accounts. Reasons for sharing varied, with “helping others” and “seeking help and social support” being the most popular (66.9% and 51.8% respectively). On the other hand, about half of the participants (49%) preferred not to share their HRI because of the “different” people among their OSN contacts, whom they did not want to share HRI with. Most participants (about 93%) considered the HRI *type* and its *recipient* to be the most significant factors affecting the magnitude of the privacy risks. Participants perceived lower privacy risks when sharing HRI with “select individual(s)” or

TABLE I
HRI CATEGORIES AND EXAMPLES.

Category	HRI Examples
Healthy living	dietary and healthy eating, physical exercise, mental exercise, emotional exercise, environmental hazards
Own experience	experience with: previous surgeries, treatments and their side effects, symptoms, specific doctor or medical institute, health insurance companies and plans
Useful found information	books, articles, websites
Mental and emotional health conditions	sad, stressed, happy, excited, depressed, mental disorder
Physical health conditions	sick, injured, not feeling good, in good shape, tired
Medical health records	personal information and address, physician and medical institute information, insurance related information, x-rays, reports, appointments and schedule
Experience of somebody else	experience with: previous surgeries, treatments and their side effects, symptoms, specific doctor or medical institute, health insurance companies and plans
HRI of people in your custody	parents, children, and others

“select group(s),” while their risk perceptions increased with the expansion of the receiver base to include “entire contacts” and “all other OSN users” respectively.

In summary, this paper makes the following contributions:

- Reports quantitative data about OSN users’ HRI sharing practices, perceived privacy risks, and risk-mitigating behaviours.
- Shows the role of OSNs in creating new contexts for sharing HRI.
- Highlights OSN users’ privacy concerns and risk-mitigating behaviour.
- Helps in understanding existing sharing practices and main privacy concerns, and contributes toward grounding into empirical data the research on HRI sharing in OSNs.

We begin by discussing related work about HRI sharing in OSNs and its privacy related implications (Section II). Then, we present our research methodology (Section III), including design, execution and evaluation details of our online survey. Next, we report collected data about OSN users’ HRI sharing practices, privacy risk perceptions and risk-mitigating behavioural responses (Section IV). In Section VI, we discuss how the results. We conclude by discussing limitations and conclusions in Sections V and Section VII respectively.

II. RELATED WORK

There is a fine line between what to share and what to keep private. Altman [21, 22] introduced “privacy regulation theory,” which defines privacy as a dynamic process of boundary regulation in response to changes in our internal state and external conditions. People achieve better privacy by minimizing the difference between their “desired” and “achieved” (perceived) privacy levels. In order to achieve desired privacy levels, people need to maintain their privacy by developing a number of behavioural mechanisms (responses),

and then expressing their desired privacy levels to others. Although Altman’s privacy regulation theory was developed long before the OSN age, the virtual spaces created by existing OSNs are only considered as new contexts for applying privacy regulation theory [23].

The benefits of sharing HRI within social networks has received considerable attention from the research community in recent years [3, 4]. Skeels [5] studied patients with serious health issues (breast cancer), and explored their personal health information (PHI) sharing behaviours in their social networks. She used a user-centric approach to capture users’ requirements in order to design and evaluate an online interactive technology that facilitates PHI sharing, while preserving information privacy through transparent and usable interface. Although we aim at addressing similar problem (investigating HRI sharing practices), our study context is different. Skeels explored PHI sharing requirements in offline social network to inform the design of her patient-centric OSN, whereas we focus on the OSN context and the HRI sharing experience in general. Moreover, she studied patients with breast cancer only, while we target active OSN users regardless of their health conditions.

Bulgurcu [18] studied OSN users’ privacy perceptions and behavioural responses with respect to their perceived privacy threats and coping motivations. She combined “coping theory” [24] from psychology with privacy literature to identify how OSN users’ privacy threat perceptions (assessment of the consequences of using OSN features) and coping perceptions (assessment of their control over using the feature), will determine their motivation to cope with a privacy threat. She studied OSN users’ behavioural responses in the form of coping motivations and preferred behaviours (emotion-focused and problem-focused coping).² In fact, individuals may undertake different risk-mitigating behaviours in response to perceived privacy threats, and therefore, in our study, we focus on the context of HRI sharing in OSNs and explore users’ perceived privacy risks and behavioural responses respectively.

III. METHODOLOGY

We studied recent literature on HRI sharing in OSNs and the associated privacy perceptions. To the best of our knowledge, previous studies, which applied different methodologies (interviews, focus groups, and surveys), were either limited to niche demographics (e.g., teenage patients, elderly patients, mothers with children in custody) [5, 17, 19, 25, 26], or were focused on non-OSN users [3, 4, 7, 27]. Although the literature on “information sharing security and privacy” was quite rich in content, it lacked the quantitative data on the HRI sharing practices of active OSN users. Moreover, the privacy aspects of HRI sharing in OSN were not clearly presented in the literature. Therefore, with the partial knowledge gained from literature, we decided to employ an online survey, as the first

²Problem-focused coping occurs when an individual minimizes the perceived threats by targeting the source of the threat (e.g., using fake names when registering for an online service). Emotion-focused coping happens when an individual mitigates the perceived threats by underestimating the negative outcomes associated with it (e.g., thinking of the gained benefits instead of perceived threats).

step toward building a better understanding of practices for HRI sharing by active OSN users’ and their perceptions of associated privacy risks.

We aimed at providing descriptive statistics that quantify HRI sharing practices while highlighting the main privacy-related issues. As a result, we surveyed active OSN users about their HRI sharing practices with their OSN peers. We also collected data about participants’ perceived risks when sharing different HRI with various user groups. Finally, we collected OSN users’ risk-mitigating behaviours at different risk levels.

A. Online Survey

We constructed an online survey³ that consisted of four sections. In the first section, we collected participants’ demographics along with information about their OSN accounts and usage. In the second section, participants reported how often they shared HRI, and identified reasons for and against sharing HRI in OSNs. Next, we asked participants to specify their perceived privacy risk levels when sharing different types of HRI with different users. In the fourth section, we asked participants to specify their preferred risk-mitigating behavioural response(s) at different levels of perceived privacy risks. Survey questions and response options are provided in Appendix A.

1) *General Usage*: We asked participants about the total number of active OSN accounts they maintained, and, for each active OSN account, we asked the following: (i) name of the OSN; (2) how often the OSN account was used; and (3) the preferred device for connecting to that OSN account (e.g., smartphone, laptop, tablet).

2) *HRI Sharing Practices*: We grouped a superset of HRI items found in the literature [5–7] into 8 unique HRI categories, as shown in Table I. For each category, we asked participants to rate the likelihood of sharing information from “never” to “always” on a 5-point Likert scale, to differentiate users who shared HRI from those who never did. We also asked participants to specify all reasons that would lead them to share and not share HRI, by selecting all that applied in the aggregated list of possible reasons.

3) *Perceived Privacy Risks*: To better understand users’ privacy risk perceptions, we asked participants to identify how much each factor (HRI recipients, HRI category, used OSNs, and current health conditions) contributes to their perceived privacy risks (responses varied on a 5-point Likert scale from “does not affect at all” to “strongly affects”). Then, we asked participants to rate their perceived privacy risks of sharing information from each HRI category on a 5-point Likert scale, from “very low or no risks” to “extreme risks,” when sharing with different users (select individuals, select groups, entire contacts, and all other OSN users).

4) *Risk-Mitigating Behaviours*: In the final section of the survey, we assumed different privacy risk levels (“very low or no risks” to “extreme risks”), and asked participants to indicate

their preferred risk-mitigating behaviour(s) when sharing HRI in OSNs. We wanted to capture the preferred behavioural responses of participants in the presence of different risk levels, regardless of the perceived benefits, and describe their risk-mitigating responses in terms of risk avoidance (not sharing) and coping behaviours (coping with threats by minimizing perceived risks).

B. Participant Recruitment

We recruited Amazon MTurk [20] workers through CrowdFlower [28], a crowdsourcing website that provides different labour channels. MTurk workers, who were shown to be more educated and younger than the general population, can be considered as a reliable source of high-quality data for research involving human-subjects [29]. CrowdFlower was used to recruit “trusted” MTurk workers, who are identified by CrowdFlower considering different factors such as their prior completed jobs and reported activities. CrowdFlower also keeps track of users’ IP addresses and used aliases (if any), and prevents multiple submissions from the same user upon request. Therefore, we used available features to limit participants to a single submission only. The posted recruitment job title, “An online survey of health related information sharing in online social networks,” did not mention privacy, while in the instructions, participants were requested to complete an online survey about their HRI sharing practices, perceived privacy risks, and corresponding behavioural responses that would take less than 30 minutes of their time. Participants were compensated with \$1 (US) through CrowdFlower for successfully completing the job.

The survey was conducted completely online, with participants’ identities anonymized throughout the study. We minimized risk to participants by anonymizing submissions and excluding any personally identifiable information from generated results and published reports. As a requirement of UBC’s Behavioural Research Ethics Boards (BREB) for conducting user studies, the recruiting material specified that participants should be 19 years of age and older. The study also required recruiting participants with at least one active OSN profile that they use regularly. After obtaining UBC’s BREB approval for conducting our study, we started the data collection process by posting a job request on CrowdFlower. Participants were provided with a link to the survey website, where they reviewed a consent form describing the research, data collection and storage policies. To guarantee compensation by CrowdFlower, participants were required to prove survey submission by entering a unique 7-digit alpha-numeric code that was assigned to them by the survey website upon survey completion.

C. Categorization of HRI Examples

Prior to conducting our survey, we reviewed the literature and found more than 30 different examples of HRI that people share online [5, 30]. They include but not limited to the following: mental or physical conditions, healthy eating and dietary information, daily exercise routines, experience with a specific doctor, previous surgeries, symptoms and side

³We administered the survey on Enterprise Feedback Management (EFM), a Canadian-hosted survey solution complying with the Freedom of Information and Protection of Privacy Act of British Columbia, where our university is located. This enterprise-level survey tool is web-based and is a comparable alternative to the US-based Survey Monkey.

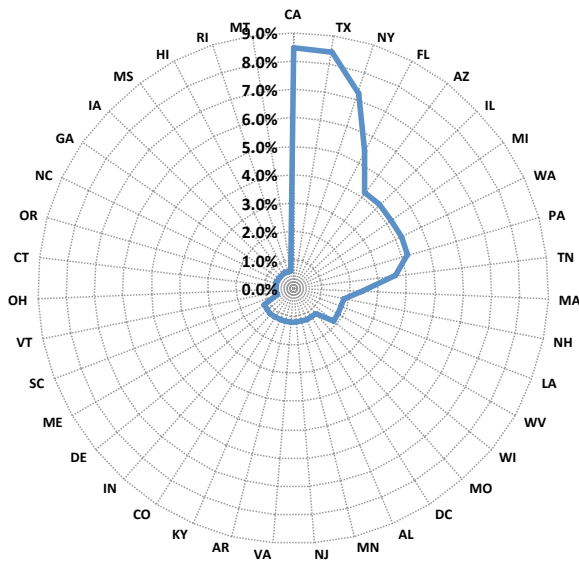


Fig. 1. Participants distribution among 39 states.

effects, insurance company information, useful online articles, information about children and elderly family members. One of the paper authors categorized the identified examples of HRI sharing according to their similarities and characteristics, as shown in Table I. To triangulate his categorization, we used an online card sorting tool *OptimalSort*⁴ to run a closed card sorting exercise, where we asked 11 participants, mostly friends and colleagues, to group different HRI examples into predefined 8 categories. Seven participants grouped the majority of the HRI examples into categories similar to our predefined ones.

IV. RESULTS

On April 4th, 2013, 191 people successfully completed our survey. Overall, the vast majority of the participants have shared some HRI through their OSN accounts. While participants showed more concerns regarding the shared HRI type and its recipient(s), they adopted variable behavioural responses to mitigate different levels of risk. In this section, we first report participant demographics and their OSN usage specifications. Then, we present results concerning active OSN users' HRI sharing practices, regularity, and reasons. Next, we present participants' perceived privacy risks when sharing HRI with different users. Finally, we report participants' preferred behavioural responses when sensing different privacy risk levels.

A. Participant Demographics and OSN Usage

Out of the 191 participants, two did not consent to the study, and therefore were removed from further analysis. The majority of the remaining 189 participants were from the US (87.8%), followed by India (9.5%). We also received a single submission from participants residing in each of the following

⁴www.optimalworkshop.com

TABLE II
PARTICIPANT DEMOGRAPHICS (166 ACTIVE OSN USERS FROM THE US).

Gender	49.4%	Male
	50.6%	Female
Age	All	19–70 ($\mu=30.4$, $\sigma=10$, median=28)
	59.0%	19–29
	24.7%	30–39
	8.4%	40–49
Completed education	7.9%	50+
	25.3%	high school
	12.7%	post-secondary diploma
	39.8%	undergraduate degree
	6.6%	community college
Employment category	10.2%	graduate degree (PhD or Masters)
	5.4%	other
	More than 18 different categories: IT (10.6%), Education (9%), Medical (7.9%), Banking (7.4%), and other categories including Student and Self-employed	

countries: Australia, Philippines, Romania, Canada, and UK. Since users' privacy risk perceptions and behaviour responses are highly dependent on the social environment and cultural context [22], we discarded submissions made by participants residing outside of the US in order to minimize the effect of cultural differences in our analysis. We report the rest of our analysis on this sample of 166 participants from the USA. They represent users from at least 39 different states (24 participants did not specify their state), with California and Texas having the highest participation rates (8.4%), followed by New York and Arizona with 7.2% and 5.4% respectively (see Figure 1).

With approximately equal number of male (49.4%) and female (50.6%) participants, our sample had ages ranging between 19 and 70 ($\mu = 30.4$, $\sigma = 10$). As Table II shows, 59% of all participants were between 19 and 29 years old, with the majority (83.7%) being between 19 and 39. Approximately 40% had completed an undergraduate degree, while the rest had completed high school (25.3%), post-secondary (12.7%), and graduate degrees (10.2%). Our sample covered over 18 employment categories, with IT (10.6%), Education (9%), Medical (7.9%), and Banking (7.4%) being most popular.

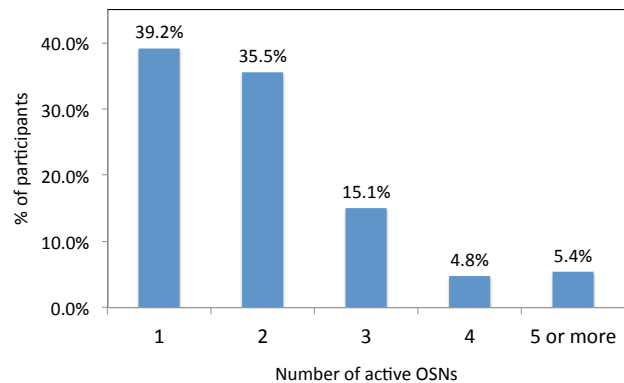


Fig. 2. Distribution of participants over the number of active OSN accounts.

Just over 39% of all participants indicated having only one active OSN account, with the rest using 2 or more. See

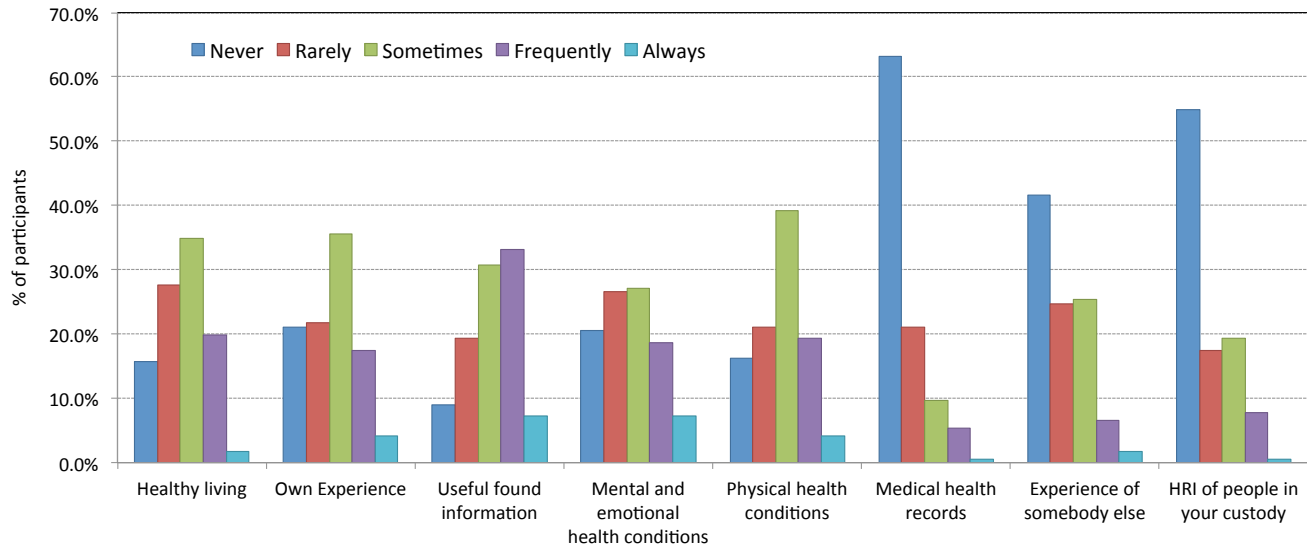


Fig. 3. Frequency of sharing HRI in OSNs.

Figure 2 for details. Out of the 140 participants who reported using their OSN account(s) every day, 135 (96.4%) were active Facebook users. We also asked about the preferred device for connecting to each OSN account (PC/Laptop, Tablet, or Smartphone). About 70% preferred to use a PC/Laptop and approximately 24% used their smartphones.

B. HRI Sharing Practices

To investigate sharing practices, we presented participants with eight HRI categories (Table I), and asked them to indicate on a 5-point Likert scale how regularly they shared information in each category. As shown in Figure 3, and summarized in Table III, 63.3% of participants never shared information about their medical health records (e.g., x-rays, treatments, name and address, insurance policy number). More than half of participants (54.8%) never shared HRI of people in their custody (e.g., children and elderly parents), while 41.6% indicated never sharing other people’s health related experience. Only 7 participants never shared any HRI in OSNs, while the rest (95.8%) indicated sharing.

To answer the question “*why do you share HRI online?*” participants selected all that applied from a list of common reasons obtained from the literature [5–7]. The provided list of

TABLE III
HRI CATEGORIES AND % OF PARTICIPANTS WHO NEVER SHARED HRI FROM EACH CATEGORY. THE THIRD COLUMN REPRESENTS THE MEDIAN RANKING FROM THE LIKERT SCALE

HRI categories	Participants	Median
Medical health records	63.3%	1
HRI of people in your custody	54.8%	1
Experience of somebody else	41.6%	2
Own Experience	21.1%	3
Mental and emotional health conditions	20.5%	3
Physical health conditions	16.3%	3
Healthy living	15.7%	3
Useful found information	9.0%	3

reasons, both *for* and *against* sharing HRI (Tables IV and V), represent general sharing motivation and are not related to any specific HRI category. As shown in Table IV, more than half of participants (66.9%) shared their health related knowledge and experience to “help others” in their social network, followed by “getting beneficial feedback” and “social support” as other reasons for sharing HRI online (51.8% and 48.2% respectively). On the other hand, for the HRI that the participants did not share in their OSNs, approximately half of them (49.4%) indicated not sharing HRI due to the existence of different people among their OSN contacts, while 43.4% of all participants considered their HRI to be personal and did not want to share it with anyone (Table V). Since we do not have exact information about users’ actual behaviour, we can not assume that privacy-related issues were the only motivation preventing participants from sharing HRI online. However, the results indicate that about half of the participants did not share HRI in their OSN(s) due to its personal nature and due to the existence of different people in their social contacts, which might be an indication of privacy concerns.

TABLE IV
REASONS FOR SHARING HRI ONLINE.

Reasons for sharing HRI	Participants
Help others by sharing personal experience and knowledge	66.9%
Seek help or social support	51.8%
Get benefits by receiving useful feedback from on-line contacts	48.2%
Seek online interactions and make discussions	44.0%
Alleviate anxiety (sharing HRI makes me feel better and less stressed)	38.6%
Promote healthy living	37.3%
Other reasons	6.6%

C. Perceived Privacy Risks

The subjective perception of privacy risks by technology users is highly context dependent [23]. In the context of HRI

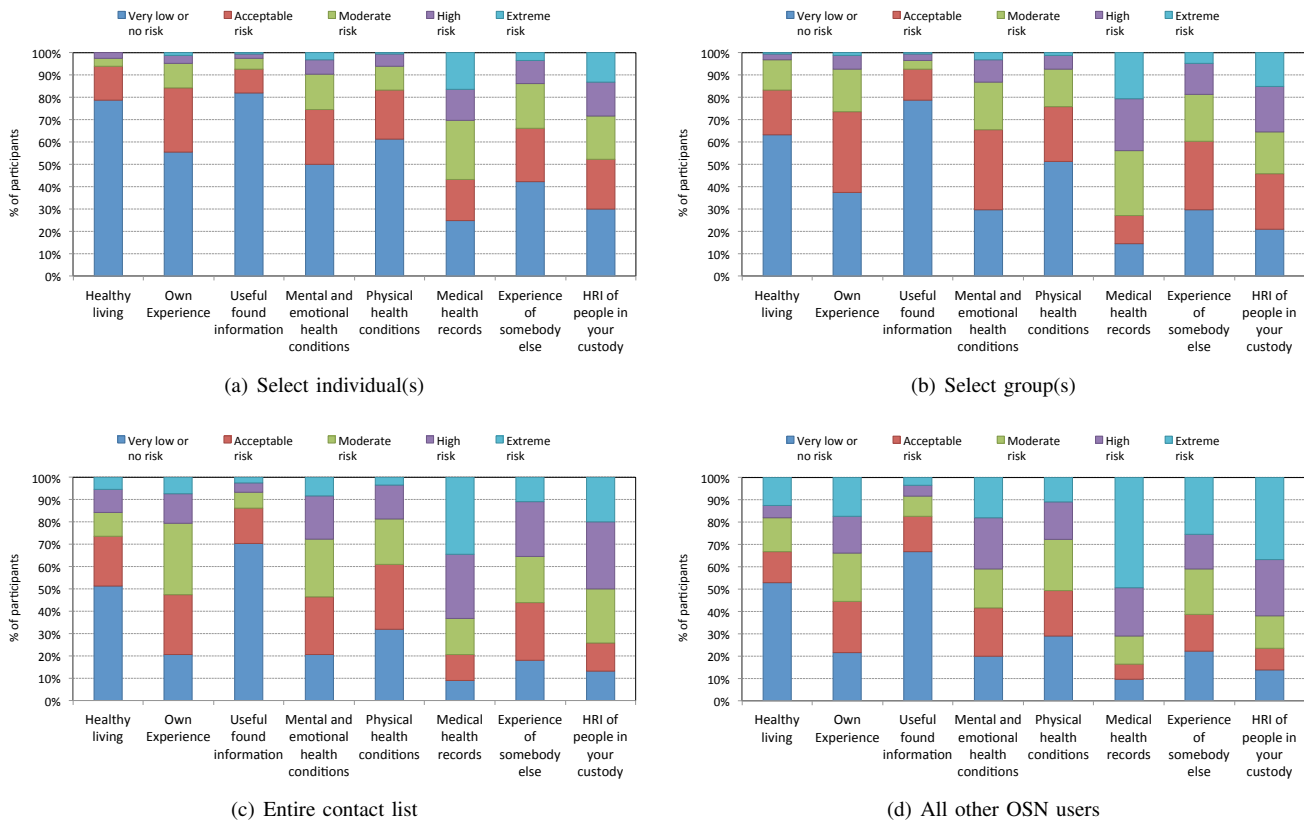


Fig. 4. Perceived privacy risks when sharing HRI with different user categories.

TABLE V
REASONS AGAINST SHARING HRI ONLINE.

Reasons against sharing HRI	Participants
I have different people in my online contact list and I prefer not to share my HRI with all of them	49.4%
My HRI is personal and I do not share it with anyone	43.4%
I do not want to be treated as “the sick” person by my online contacts	39.2%
I do not want my online contacts to worry about me by receiving bad news about my health	34.9%
I do not want my online contacts to know about my HRI	34.9%
I prefer to share my HRI with my doctor	30.1%
I prefer to share my HRI offline	22.3%
My online contacts are not interested in my HRI	20.5%
Other reasons	4.8%

sharing in OSNs, the key factors are as follows: (1) recipient of information; (2) HRI type and category; (3) the OSN where information is shared; and (4) the mental and physical health conditions of the individual at time of sharing. We asked participants to rate on a 5-point Likert scale how much each of the above four factors would affect their perceived privacy risks (responses varied between “does not affect” to “strongly affects”). While more than 85% of participants indicated that all four factors affect their perceived privacy risks, results showed that the HRI “recipient” and its “type and category,” *strongly affected* the perceived privacy risks of approximately 30% of the participants.

Using results from our literature review on the effect of the recipient and the type of information on the perceived privacy

risks, when sharing information in an OSN, we asked users to rate their perceived privacy risks, when sharing information from each HRI category with different user groups. We used a 5-point Likert scale (with responses varying from “very low or no risk” to “extreme risk”) and the following user groups: select individual(s), select group(s), entire contact list, and all other OSN users. Although it was not explicitly mentioned to the participants, select individual(s) and group(s) implied trusted user categories. We tried to avoid using the word “trust” in order to minimize response bias, while implying trust by presenting selective user choice when sharing HRI. Our aim was to capture users’ perceived privacy risk variations with respect to the HRI category and the number of trusted recipients, and compare it with the perceived privacy risks of sharing with other user groups that possibly include untrusted users such as the participant’s entire contact list or all other users in the OSN. Figure 4 shows the results.

More than 66.9% of participants perceived *very low or no risks* when sharing “useful found information” with their “entire contacts” or “all other OSN users” (perceived privacy risks of sharing with select individual(s) and group(s) was 81.9% and 78.9% respectively). This low sensitivity toward sharing “useful found information” might be due to the nature of that type of HRI, which might not directly relate to a person’s health. For other HRI categories such as “healthy living,” “physical health conditions,” “own experience,” and “mental and emotional health conditions,” the percentage of participants who perceived sharing HRI to be of “very low or

no risks” decreased with respect to the extent in the recipient base. Interestingly, as shown in Figure 4, the difference in the number of participants who perceived sharing the eight HRI categories to be of “very low or no risks” was significantly smaller, when compared to the sharing HRI with “entire contact list” and “all other OSN users”. This might indicate that a portion of participants perceive privacy risks related to sharing HRI with their entire contact list to be similar to sharing with all other OSN users. From a different perspective, about 16% of participants perceived sharing information about their “medical health records” to be of *extreme risk*, even if shared with select individuals. Similarly, sharing “HRI of people in custody” raised high concerns by participants, with over 13% of participants perceiving *extreme* privacy risks, even when sharing with select individuals. As shown in Figure 4, participants’ risk perceptions significantly increased when the recipients of information expanded from “select individuals” to “all other OSN users”.

To find out if there is a significant difference in participants’ perceived privacy risks when sharing HRI with different user groups, we performed the Friedman rank sum tests over all participants’ responses. The Friedman test is a non-parametric statistical test similar to the parametric repeated measures ANOVA, which is used when the dependent variables are ordinal (e.g., Likert scale outcomes), and where the behaviours of the same sample of subjects are assessed for existing differences under differing conditions. For all HRI categories, the resulted p-values were sufficiently small ($p < 0.05$), which suggested that somewhere in the data, there existed a significant difference between participants’ perceived privacy risks, when sharing HRI with different user groups.

To identify the actual location of the significant difference, the Friedman test was followed with a series of post hoc Wilcoxon tests on each combination of user categories. For each pair of categories, we tested a separate null hypothesis that there was no significant change in the attitude, when sharing every HRI between “select individual(s)” and “select group(s),” “select individual(s)” and “entire contacts,” “select individual(s)” and “all other OSN users,” and so on for all 6 combinations of user groups. To reduce the chances of obtaining false-positive results (type I errors), when multiple pair-wise tests were performed on a single set of data, we applied the Bonferroni adjustment ($\alpha_{new} = 0.0083$)⁵ to a series of post hoc Wilcoxon matched pair tests. The tests resulted in no statistically significant difference ($p > 0.0083$) in participants’ perceived privacy risks, when sharing “healthy living” and “useful found information” categories between “entire contacts” and “all other OSN users,” as shown in Tables VI and VIII. Moreover, no statistically significant difference was found in participants’ perceived privacy risks when sharing “useful found information” between “select individual(s)” and “select group(s)” (Table VIII). For the rest of the pair-wise comparisons, we found statistically significant difference ($p < 0.0083$) in participants’ perceived privacy risks, when sharing HRI of different categories between different

user groups (see Appendix B).

D. Risk Mitigation and Behavioural Responses

We used risk avoidance and coping theory [18] to collect and analyze participants’ responses about mitigating risks. We assumed that the participants could avoid privacy risks by not sharing HRI (risks avoidance), or they could cope with possible privacy threats by manipulating the source of the risks (the shared HRI, its recipients, and the choice of the OSN). We also assumed that the participants were always able to share HRI without taking any preemptive risk-mitigating actions.

In general, participants reported a diverse range of behaviour, when exposed to different risk levels (from “very low or no risks” to “extreme risk”), with just over 58% choosing not to share HRI at extreme risk levels. At “acceptable privacy risk” levels, which appeared 2nd in the Likert scale right after “very low or no risks,” about 50% of the participants preferred to cope with privacy threats by manipulating shared HRI and/or filtering its recipients.

V. LIMITATIONS AND FUTURE WORK

There exist several limitations that relate to the participants recruitment and the survey questions. First, the participants had different background, knowledge and HRI sharing experience, which increased the subjectivity of their responses in terms of perceived privacy risks and behavioural responses. Second, as with any survey, the answers provided were self-reported. Finally, the survey questions were based on the authors’ knowledge and understanding of the literature.

To minimize threats to internal validity, we used an online crowdsourcing tool (CrowdFlower) to recruit MTurk workers, who represent a participant pool with different demographics and knowledge levels, resulting in a sample similar or sometimes better than participant samples recruited using conventional methods (emails, flyers, online posts, and ads) [29, 31]. Our recruited participants were from 39 different states in the US, with ages between 19 and 70 years, and almost equal male and female ratios. We informed the design of our survey questions and responses by findings from literature [3, 5–7, 18]. Moreover, we provided participants with open-ended response choices to express their answers in their own words, when necessary. Finally, the survey was pre-tested and reviewed with a number of colleagues and co-workers with expertise in conducting user studies in order to achieve consistent understanding of the survey questions and the response choices.

Although our participants sample covered most states in the US (39 out of 50), and included participants from various age ranges, education levels, and employment categories, the generalizability (external validity) of our results could be hampered by the limited representativeness of the participants. Also, the majority of the participants represented Facebook users (about 96%), while recent statistics showed that about 67% of online US adults are Facebook users [7]. However, the aim of this study was to explore current users’ behaviour in order to identify future research directions and therefore,

⁵Considering the Bonferroni adjustment, α is divided by the number of pair-wise combinations in the user categories ($\alpha_{new} = 0.05/6$).

results generalizability might not be the primary outcome of the study.

In retrospect, we realize that our survey questions were at times too abstract and did not contain specific scenarios. In the future, by creating more realistic scenarios for sharing HRI in OSNs, and by placing participants in the context of sharing HRI in OSNs, one can capture users' behaviour in more details, resulting in more accurate findings with respect to risk perceptions and behavioural responses. Finally, prior studies showed that the existence of other factors such as benefits related to sharing HRI, could be considered as a motive for shaping users' perceptions and forming their behaviour, and therefore, we would like to incorporate those factors in our scenarios to understand how they may affect users' perceived risks and risk-mitigating behavioural responses.

VI. DISCUSSION

Our results contribute to better understanding of HRI practices in OSN, perceived privacy risks, and preferred risk-mitigating behaviour. We found that most active OSN users (about 96%), who indicated using their OSN account(s) on a daily basis, had shared some HRI in the past. While the majority of participants shared some HRI, more than half of them never shared information about their medical health records and HRI of people in their custody. On the one hand, there is a caveat here that our participants might have never shared such information, if they did not have access to it in the first place. For example, the participants might not have access to electronic health records, or simply did not have any dependents or people in their custody, and therefore, could not share such HRI online. On the other hand, the results show that even when sharing with select individuals, sharing medical health records and information about people in custody were perceived as "high" or "extreme" risks by approximately 30% of the participants.

After performing the Friedman and the post hoc Wilcoxon tests, the effect of "information recipient" and "HRI type and category" factors were clearly observed in increasing perceived privacy risks associated with sharing HRI in OSNs (Section IV-C). As shown in Figure 4, for almost all HRI categories, the perceived privacy risks of sharing HRI in OSNs increased toward higher risk levels, with respect to the expansion of the recipient base from "select individual(s)" and "select groups(s)" to include more recipients (i.e., "entire contacts" and "all other OSN users"). For HRI categories of "healthy living" and "useful found information," the participants expressed their perceived privacy risks with lower granularity, where they indicated sharing HRI among different user group pairs with similar perceived risks (Tables VI and VIII in Appendix B). Therefore, while different recipient groups affected the perceived privacy risk outcomes, the differences in the type of HRI were also shown to influence the results of the perceived privacy risks.

Medical health records may include personally identifiable information along with information that is directly related to an individual and her health, and therefore, participants' concerns regarding sharing such information is justifiable. However,

HRI of people in custody has received considerable attention by participants as well. We believe that people handle HRI of other individuals in their custody with extra care and with a sense of responsibility toward protecting them against privacy threats. Moreover, the category "HRI of people in custody" represents an abstract view of all HRI related to a person in custody, including their medical health records. Therefore, participants might have considered an overall risk perception with respect to the most sensitive HRI categories, such as medical records of people in their custody, and as a result, indicated high privacy concerns when sharing such HRI.

We explored risk-mitigating behaviour of participants at different risk levels. We found that even at extreme risk levels, over 41% of the participants shared HRI, while minimizing risks using different techniques (manipulate shared HRI, filter recipients, and change used OSN). Furthermore, at "acceptable risk levels," which represents the lowest assumed risk level after "very low or no risks," more than half of participants preferred to mitigate privacy risks by manipulating and filtering shared HRI and its recipients respectively. Meanwhile, about 19% of participants were extra cautious and preferred to avoid risks by not sharing HRI even at "acceptable risk levels". In 2012, a study of Europeans' perception, behaviour, and attitude toward sharing HRI identified three sharing profiles: self-revealing (aware of risks but reveals information to gain benefits), indifferent, and cautious [3]. We believe that our participants' behaviour could be profiled similarly, classifying participants who shared HRI regardless of the perceived risk levels as "indifferent," whereas "self-revealing" participants are those who were motivated to share their HRI, but preferred to minimize risks by adopting available methods, and finally, "cautious" participants, who refused to share HRI under different circumstances.

VII. CONCLUSION

Our survey contributes to understanding HRI sharing practices, risk perceptions and risk-mitigating behavioural responses. We quantified the HRI sharing practices in OSNs by showing that almost 96% of OSN users shared some HRI with their social peers. We found that participants were more sensitive toward sharing their medical health records, while considering sharing HRI of people in their custody with extra care. We investigated OSN users' risk-mitigating behaviour and found that some people (indifferent and cautious) tend to follow unique behaviour regardless of perceived risk levels, while others (self-revealing) behaved differently at the presence of variable risk levels.

Previous studies were limited to niche user categories (patients, teenage OSN users, elderly people, etc.), and they did not focus on the actual behaviour and concerns of active OSN users, when it comes to HRI sharing. By providing quantitative evidence showing that HRI sharing is in fact a new application of existing OSNs, along with quantifying OSN users' privacy concerns and risk-mitigating behaviour, we are contributing towards building a solid ground for future studies that will address more specific questions.

ACKNOWLEDGEMENTS

We would like to thank all LERSSE members (Yazan Boshmaf, Pooya Jaferian, Hootan Rashtian, SanTsai Sun and Ildar Muslukhov) for their kind help, advice and feedback. We also would like to thank Hyounghick Kim for his guidance and feedback during early stages of the project. This research is partially supported through funding from the NSERC Inter-networked Systems Security Network (ISSNet).

REFERENCES

- [1] A. Mueller, J. Pearson, C. Muller, K. Frank, and A. Turner, "Sizing up peers: Adolescent girls' weight control and social comparison in the school context," *Journal of Health and Social Behavior*, vol. 51, no. 1, pp. 64–78, March 2010.
- [2] J. T. Kullgren, A. B. Troxel, G. Loewenstein, D. A. Asch, L. A. Norton, L. Wesby, Y. Tao, J. Zhu, and K. G. Volpp, "Individual- versus group-based financial incentives for weight loss: A randomized, controlled trial," *Annals of Internal Medicine*, vol. 158, no. 7, pp. 505–514, 2013. [Online]. Available: <http://dx.doi.org/10.7326/0003-4819-158-7-201304020-00002>
- [3] F. Lupiáñez-villanueva, W. Lusoli, M. Bacigalupo, I. Maghiros, N. Andrade, and C. Codagnone, "Health-related information as personal data in Europe: Results from a representative survey in Eu27," *Medicine 2.0: Ethical and legal issues, confidentiality and privacy*, 2012.
- [4] D. M. Zulman, K. M. Nazi, C. L. Turvey, T. H. Wagner, S. S. Woods, and L. C. An, "Patient interest in sharing personal health record information: A web-based survey," *Annals of Internal Medicine*, vol. 155, no. 12, pp. 805–810, 2011. [Online]. Available: <http://dx.doi.org/10.1059/0003-4819-155-12-201112200-00002>
- [5] M. M. Skeels, "Sharing by design: Understanding and supporting personal health information sharing and collaboration within social networks," Ph.D. dissertation, University of Washington, 2010.
- [6] S. Fox, "The social life of health information," Pew Research Center Report, May 2011.
- [7] S. Fox and M. Duggan, "Health online 2013," Pew Research Center's Internet and American Life Project, January 2013.
- [8] J. Becker and H. Chen, "Measuring Privacy Risk in Online Social Networks," in *Proceedings of W2SP 2009: Web 2.0 Security and Privacy*, May 2009. [Online]. Available: <http://w2spconf.com/2009/papers/s2p2.pdf>
- [9] B. Krishnamurthy and C. E. Wills, "On the leakage of personally identifiable information via online social networks," in *WOSN '09: Proceedings of the 2nd ACM Workshop on Online Social Networks*. New York, NY, USA: ACM, 2009, pp. 7–12.
- [10] J. A. Calandrino, A. Kilzer, A. Narayanan, E. W. Felten, and V. Shmatikov, "'you might also like:': privacy risks of collaborative filtering," in *Proceedings of the 2011 IEEE Symposium on Security and Privacy*, ser. SP '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 231–246. [Online]. Available: <http://dx.doi.org/10.1109/SP.2011.40>
- [11] R. Gross and A. Acquisti, "Information revelation and privacy in online social networks," in *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*, ser. WPES '05. New York, NY, USA: ACM, 2005, pp. 71–80. [Online]. Available: <http://doi.acm.org/10.1145/1102199.1102214>
- [12] P. V. Eecke and M. Truysens, "Privacy and social networks," *Computer Law and Security Review*, vol. 26, no. 5, pp. 535 – 546, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0267364910001093>
- [13] S. Campbell, "The privacy risks of sharing health info online," Reputation.com.
- [14] D. Masys, "Of codes, genomes, and electronic health records: It's only sensitive if it hurts when you touch it," Usenix HealthSec Workshop, August 2012.
- [15] B. McGowan. (2012) How much of your personal health information should be online? [Online]. Available: <http://www.theatlantic.com/health/archive/2012/03/howmuchofyourpersonalhealthinformationshouldbeonline/254375/>
- [16] J. Williams, "Social networking applications in health care: threats to the privacy and security of health information," in *Proceedings of the 2010 ICSE Workshop on Software Engineering in Health Care*, ser. SEHC '10. New York, NY, USA: ACM, 2010, pp. 39–49. [Online]. Available: <http://doi.acm.org/10.1145/1809085.1809091>
- [17] M. Velden and K. El Emam, "'not all my friends need to know': a qualitative study of teenage patients, privacy, and social media," *Journal of the American Medical Informatics Association*, vol. 20, no. 1, pp. 16–24, July 2012.
- [18] B. Bulgurcu, "Understanding the information privacy-related perceptions and behaviors of an online social network user," Ph.D. dissertation, The Faculty of Graduate Studies, Business Administration, University of British Columbia, August 2012.
- [19] "How U.S. moms share and spread HEALTH information," www.lucidmarketing.com, 2011.
- [20] "Amazon Mechanical Turk," <https://www.mturk.com/>, 2005.
- [21] I. Altman, *The environment and social behavior*. Brooks/Cole Monterey, CA, 1975.
- [22] —, "Privacy regulation: culturally universal or culturally specific?" *Journal of Social Issues*, vol. 33, no. 3, pp. 66–84, 1977.
- [23] L. Palen and P. Dourish, "Unpacking 'privacy' for a networked world," in *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM, 2003, pp. 129–136.
- [24] R. Lazarus, "Coping theory and research: Past, present, and future," *Fifty Years of the Research and Theory of RS Lazarus: An Analysis of Historical and Perennial Issues*, pp. 366–388, 1993.
- [25] P. Ressler, Y. Bradshaw, L. Gualtieri, and K. Chui, "Communicating the experience of chronic pain and

illness through blogging,” *Journal of Medical Internet Research*, vol. 14, no. 5, p. e143, 2012.

- [26] A. Hartzler, M. Skeels, M. Mukai, C. Powell, P. Klasnja, and W. Pratt, “Sharing is caring, but not error free: Transparency of granular controls for sharing personal health information in social networks,” in *AMIA Annual Symposium*, vol. 2011, 2011, pp. 559–568.
- [27] T. W. Feeley and K. I. Shine, “Access to the medical record for patients and involved providers: Transparency through electronic tools,” *Annals of Internal Medicine*, vol. 155, no. 12, pp. 853–854, 2011. [Online]. Available: [+http://dx.doi.org/10.1059/0003-4819-155-12-201112200-00010](http://dx.doi.org/10.1059/0003-4819-155-12-201112200-00010)
- [28] “CrowdFlower,” www.crowdfunder.com, 2007.
- [29] B. Ur, P. G. Kelley, S. Komanduri, J. Lee, M. Maass, M. L. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer *et al.*, “How does your password measure up? the effect of strength meters on password creation,” in *Proc. USENIX Security*, 2012.
- [30] S. Fox, “Health topics,” Pew Research Center Report, February 2011.
- [31] A. J. Berinsky, G. A. Huber, and G. S. Lenz, “Evaluating online labor markets for experimental research: Amazon.com’s mechanical turk,” *Political Analysis*, vol. 20, no. 3, pp. 351–368, 2012.

APPENDIX A SURVEY QUESTIONS

A. Questions About Participants’ Demographics

- 1) Gender:
 - Male
 - Female
- 2) How old are you?
(19 to 99) years old
- 3) What is your highest level of completed education?
 - Less than high school
 - High school (secondary school)
 - Diploma (post-secondary courses)
 - Undergraduate university degree (Bachelor’s)
 - Graduate university degree (Masters’s or PhD)
 - Community college or professional school (College degree)
 - Other (please specify)
- 4) Employment category: [Select a category that best fits your current job]
 - A range of employment categories (e.g., Education, Business, Engineering, ...), or
 - Other (please specify)
- 5) Current country of residence?
 - Select from list, or
 - Other (please specify)

B. Questions About OSN Usage

- 1) How many active OSN profiles do you maintain? [Active accounts are those which you regularly use to connect to people, share information, and perform online activities]
 - 1
 - 2
 - 3
 - 4
 - 5 or more
- 2) For your most frequently used OSN account(s), provide the name of the OSN and specify how regularly do you log into your account. [Information for at least one OSN is required, you may skip the rest if does not apply]
The following questions will be repeated for each specified active OSN account:
 - OSN name [Select from list, or specify other OSN names]
 - How regularly do you log into your above OSN account?
 - Daily
 - Weekly
 - Monthly
 - Every few months
 - Other
 - Select the device that you use the most to connect to the specified OSN account.
 - Personal Computer/Laptop/Desktop
 - Tablet (e.g., iPad)
 - Smartphone (e.g., iPhone)
 - Other (please specify)

C. Questions About HRI Sharing Behaviours in OSNs

- 1) For each HRI category, how often do you share information in OSNs? [The survey presented 8 HRI categories with corresponding responses in the form of a 5-point Likert scale with anchor points specified as: never, sometimes, and always]
 - Refer to Table I for HRI categories.
- 2) Specify other types of HRI that you share and do not appear in the above list.
- 3) Why do you share your HRI in OSNs? [Select all that apply]
 - Help others by sharing personal experience and knowledge
 - Get benefits by receiving useful feedback from online contacts
 - Seek help or social support
 - Alleviate anxiety (Sharing HRI makes me feel better and less stressed)
 - Seek online interactions and make discussions
 - Promote healthy living
 - Other (please specify)
- 4) For HRI that you do not share in OSNs, describe your choice by selecting all reasons that apply from below.
 - I do not want to be treated as “the sick” person by my online contacts
 - I do not want my online contacts to worry about me by receiving bad news about my health
 - I have different people in my online contact list and I prefer not to share my HRI with all of them
 - My HRI is personal and I do not share it with anyone
 - I prefer to share my HRI with my doctor
 - I do not want my online contacts to know about my HRI
 - My online contacts are not interested in my HRI
 - I prefer to share my HRI offline
 - Other (please specify)

D. Questions About Privacy Risk Perceptions and Behavioural Responses

- 1) When sharing HRI in OSNs, how much each of the following factors contribute to the increase of your perceived privacy risks: [Responses vary on a 5-point Likert scale with anchor points specified as: does not affect, slightly affects, and strongly affects]
 - The receiver(s) of the HRI
 - HRI type and category
 - The OSN platform where the HRI is shared
 - Your current physical or mental health condition (e.g., depressed, healthy, suffer from chronic pain, etc.)
- 2) For every HRI category (Table I), how do you evaluate your perceived privacy risk level when sharing HRI with the given four user categories:
 - Select individual(s)
 - Select group(s)

- Your entire contact list
- All other OSN users

[Select the perceived privacy risk level from a 5-point Likert scale with anchor points: very low (or none), moderate, and extreme]

- 3) “Behavioural responses” refer to different actions undertaken by an individual to mitigate the perceived privacy risks. Assume you want to share some HRI in OSN: For each assumed privacy risk level (very low, acceptable, moderate, high, extreme), what would be your preferred behavioural response(s)? [select all that apply for each risk level]
 - Avoid risks by not sharing HRI
 - Minimize risks by manipulating shared HRI (e.g., change or remove personal and identifiable information)
 - Minimize risks by filtering recipients (e.g., share with specific individual(s) or group(s))
 - Minimize risks by changing the used OSN (e.g., share HRI in other trusted OSN)
 - Accept risks and share HRI without any preemptive actions

APPENDIX B

PERCEIVED PRIVACY RISKS: WILCOXON POST HOC TESTS

We calculated all p-values after applying a series of post hoc Wilcoxon matched pairs tests on the second question from Appendix A-D. For us to accept the alternative hypothesis, the p-value has to be less than the Bonferroni adjusted critical value, which in this instance is 0.0083. For the sake of simplicity, we only include tables that contain non-zero entries.

TABLE VI

HEALTHY LIVING: P-VALUES FOR DIFFERENT RECIPIENT COMPARISONS.

	Select individual(s)	Select group(s)	Entire contacts	All other OSN users
Select individual(s)		0.0000	0.0000	0.0000
Select group(s)			0.0000	0.0000
Entire contacts				* 0.0414
All other OSN users				

TABLE VII

OWN EXPERIENCE: P-VALUES FOR DIFFERENT RECIPIENT COMPARISONS.

	Select individual(s)	Select group(s)	Entire contacts	All other OSN users
Select individual(s)		0.0000	0.0000	0.0000
Select group(s)			0.0000	0.0000
Entire contacts				0.0004
All other OSN users				

TABLE VIII
USEFUL FOUND INFORMATION: P-VALUES FOR DIFFERENT RECIPIENT
COMPARISONS.

	Select in- dividual(s)	Select group(s)	Entire contacts	All other OSN users
Select in- dividual(s)		* 0.2836	0.0000	0.0000
Select group(s)	————		0.0005	0.0000
Entire contacts	————	————		* 0.0580
All other OSN users	————	————	————	

TABLE IX
PHYSICAL HEALTH CONDITIONS: P-VALUES FOR DIFFERENT RECIPIENT
COMPARISONS.

	Select in- dividual(s)	Select group(s)	Entire contacts	All other OSN users
Select in- dividual(s)		0.0003	0.0000	0.0000
Select group(s)	————		0.0000	0.0000
Entire contacts	————	————		0.0000
All other OSN users	————	————	————	

TABLE X
EXPERIENCE OF SOMEONE ELSE: P-VALUES FOR DIFFERENT RECIPIENT
COMPARISONS.

	Select in- dividual(s)	Select group(s)	Entire contacts	All other OSN users
Select in- dividual(s)		0.0000	0.0000	0.0000
Select group(s)	————		0.0000	0.0000
Entire contacts	————	————		0.0018
All other OSN users	————	————	————	